Science
AAAS

X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic Escherichia coli

Devapriya Choudhury, *et al.*

Science **285**, 1061 (1999);
DOI: 10.1126/science.285.5430.1061

strand and inserted into the groove of each preceding subunit. Insertion parallel to strand F yielded a rod with a star-shaped cross section, inconsistent with electron microscopy data. Insertion antiparallel to strand F produced a pilus with a helical symmetry having dimensions similar to those experimentally observed (*5*, *6*) (Fig. 4). Thus, donor strand complementation with the chaperone results in an atypical Ig fold, whereas donor strand exchange between subunits produces a canonical variable-region Ig fold in the mature pilus (*24*). Stereochemical complementarity between the NH_2-terminal motifs and grooves of the various subunits most likely restricts the order of subunit assembly. Thus, the molecular basis for the adaptor function of PapK may in part be a consequence of its NH_2-terminal motif fitting the groove of PapE and its groove accommodating the NH_2-terminal motif of PapA with stereochemical specificity.

References and Notes

1. S. J. Hultgren, C. H. Jones, S. N. Normark, in *Escherichia coli and Salmonella Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (ASM Press, Washington, DC, ed. 2, 1996), pp. 2730–2756.
2. R. A. Hull, R. E. Gill, P. Hsu, B. H. Minshew, S. Falkow, *Infect. Immun.* **33**, 933 (1981).
3. S. J. Hultgren *et al.*, *Cell* **73**, 887 (1993).
4. M. J. Kuehn, J. Heuser, S. Normark, S. J. Hultgren, *Nature* **356**, 252 (1992).
5. E. Bullitt and L. Makowski, *ibid.* **373**, 164 (1995).
6. M. Gong and L. Makowski, *J. Mol. Biol.* **228**, 735 (1992).
7. F. Jacob-Dubuisson, J. Heuser, K. Dodson, S. Normark, S. Hultgren, *EMBO J.* **12**, 837 (1993).
8. J. A. Roberts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11889 (1994).
9. B. Lund, F. Lindberg, B.-I. Marklund, S. Normark, *ibid.* **84**, 5898 (1987).
10. R. Striker, U. Nilsson, A. Stonecipher, C. Magnusson, S. J. Hultgren, *Mol. Microbiol.* **16**, 1021 (1995).
11. D. G. Thanassi, E. T. Saulino, S. J. Hultgren, *Curr. Opin. Microbiol.* **1**, 223 (1998).
12. D. L. Hung, S. D. Knight, R. M. Woods, J. S. Pinkner, S. J. Hultgren, *EMBO J.*, **15**, 3792 (1996).
13. F. Lindberg, J. M. Tennent, S. J. Hultgren, B. Lund, S. Normark, *J. Bacteriol.* **171**, 6052 (1989).
14. A. Holmgren and C.-I. Branden, *Nature* **342**, 248 (1989).
15. M. J. Kuehn, S. Normark, S. J. Hultgren, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10586 (1991).
16. E. Bullitt *et al.*, *ibid.* **93**, 12890 (1996).
17. C. H. Jones, P. N. Danese, J. S. Pinkner, T. J. Silhavy, S. J. Hultgren, *EMBO J.* **16**, 6394 (1997).
18. M. J. Kuehn *et al.*, *Science* **262**, 1234 (1993).
19. G. F. Soto *et al.*, *EMBO J.* **17**, 6155 (1998).
20. K. W. Dodson, F. Jacob-Dubuisson, R. T. Striker, S. J. Hultgren, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3670 (1993).
21. D. G. Thanassi *et al.*, *ibid.* **95**, 3146 (1998).
22. Crystals of wild-type or selenomethionine (SeMet)–containing PapD-PapK complexes (12 mg/ml) were grown by vapor diffusion using the hanging drop method against a reservoir containing 10 to 15% (w/v) polyethylene glycol (molecular weight: 6000), 100 mM potassium acetate and 200 to 400 mM sodium acetate at pH 4.6 [A. McPherson, *Eur. J. Biochem.* **189**, 23 (1990)]. Crystals were in space group $P2_12_12_1$ with cell dimensions $a = 62.1$ Å, $b = 63.6$ Å, and $c = 92.2$ Å, and with one complex in the asymmetric unit. Native and SeMet data sets to 2.7 Å and 2.5 Å resolution, respectively, were collected on single flash-cooled crystals in the laboratory setting (native and SeMet single data sets, Table 1). A SeMet PapD-PapK crystal was also used to collect multiwavelength anomalous dispersion (MAD) data at four wavelengths to a resolution of 2.4 Å at the National Synchrotron Light Source (NSLS) (SeMet–1 through –4, Table 1). All data were reduced and processed using the programs DENZO and SCALEPACK [Z. Otwinoski, in *Proceedings of the CCP4 Study Weekend*, L. Sawyers, N. Isaacs, S. Bailey, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62].
23. The structure of the PapD-PapK complex was solved using MAD phasing. The native and SeMet-single data sets were used to generate a difference Patterson map with the program HEAVY [T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr.* **A39**, 813 (1983)]. The positions of the three selenium atoms in SeMet PapD (residues 18, 66, and 172) were determined with the program HASSP [T. C. Terwilliger, S.-H. Kim, D. Eisenberg, *ibid.* **A43**, 1 (1987)] and used to calculate phases based on the MAD data with the program SHARP [E. De La Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997)]. An interpretable electron density map was obtained after solvent flipping with the program SOLOMON [J. P. Abrahams and A. G. W. Leslie, *Acta Crystallogr.* **D52**, 32 (1996)]. PapD and PapK were built into the electron density with the program O [T. A. Jones and S. Thirup, *EMBO J.* **5**, 819 (1986); T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)]; the resulting density map was of sufficient quality to unequivocally assign the sequence (Fig. 1A). The model was refined using CNS version 0.5 [A. T. Brünger *et al.*, *ibid.* **D54**, 905 (1998)] against the SeMet-3 structure factor amplitudes, using both positional and simulated annealing refinement. The final model containing 104 well-ordered water molecules has R and R_{free} values of 23.8% and 27.4%, respectively (Table 1). The model of PapD does not include residues 2 to 218 of PapD, and residues Arg^{76} and Gln^{84} in PapD were built as alanines. The model of PapK is complete but for eight residues located at the NH_2-terminus. All residues in PapA and PapD are located in either the most favored or the allowed regions of the Ramachandran plot [G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.* **23**, 283 (1968)]. Coordinates have been deposited at the Protein Data Bank (entry code 1PDK).
24. E. Y. Jones, *Curr. Opin. Struct. Biol.* **3**, 846 (1993).
25. D. Choudhury *et al.*, *Science* **285**, 1061 (1999).
26. L. N. Skanlin, J. S. Pinkner, C.-I. Branden, S. J. Hultgren, *EMBO J.* **11**, 4747 (1992).
27. M. P. Schlunegger, M. J. Bennett, D. Eisenberg, *Adv. Protein Chem.* **50**, 61 (1997).
28. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
29. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
30. A. Nicholls, K. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
31. We thank A. B. Herr for help with MAD data collection, G. Soto and D. Hung for help in preparing figures, and C. Ogata and the staff of beamline X4A at NSLS. Supported by NIH grants RO1DK51406 and RO1AI29549 (S.J.H.) and RO1CM54033 (G.W.).

25 March 1999; accepted 9 July 1999

X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *Escherichia coli*

Devapriya Choudhury,[1] Andrew Thompson,[2] Vivian Stojanoff,[3] Solomon Langermann,[4] Jerome Pinkner,[5] Scott J. Hultgren,[5]* Stefan D. Knight[1]*

Type 1 pili—adhesive fibers expressed in most members of the Enterobacteriaceae family—mediate binding to mannose receptors on host cells through the FimH adhesin. Pilus biogenesis proceeds by way of the chaperone/usher pathway. The x-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli* at 2.5 angstrom resolution reveals the basis for carbohydrate recognition and for pilus assembly. The carboxyl-terminal pilin domain of FimH has an immunoglobulin-like fold, except that the seventh strand is missing, leaving part of the hydrophobic core exposed. A donor strand complementation mechanism in which the chaperone donates a strand to complete the pilin domain explains the basis for both chaperone function and pilus biogenesis.

Type 1 pili are adhesive fibers expressed in *E. coli* as well as in most members of the Enterobacteriaceae family (*1*). They are composite structures in which a short-tip fibrillar structure containing FimG and the FimH adhesin (and possibly the minor component FimF as well) are joined to a rod composed predominantly of FimA subunits (*1*). The FimH adhesin mediates binding to mannose oligosaccharides (*2*, *3*). In uropathogenic *E. coli*, this binding event has been shown to play a critical role in bladder colonization and disease (*4*). Type 1 pilus biogenesis proceeds by way of a highly conserved chaperone/usher pathway that is involved in the assembly of over 25 adhesive organelles in Gram-negative bacteria (*5*). The usher forms an oligomeric channel in the outer membrane with a pore size of ~2.5 nm (*6*) and mediates subunit translocation across the outer membrane. Periplasmic chaperones consist of two immunoglobulin-like domains with a deep cleft between the two domains (*7*–*9*). Chaperones stabilize pilus subunits and prevent them from participating in premature interactions in the

Table 1. Summary of data collection and MAD structure determination. Two seleno-methionated FimC-FimH crystals (space group $C2$, a = 139.1 Å, b = 139.1 Å, c = 214.5 Å, β = 90.0 Å) exhibiting strong pseudo $P4_12_12$ symmetry were used to collect MAD (*23*) data on BM14 of the European Synchrotron Radiation Facility. Data were recorded at each of three wavelengths corresponding to the peak of the Se white line, the point of inflexion of the K absorption edge, and a remote wavelength by using a MAR charge-coupled device detector. Data were reduced with the program HKL2000 (*23*), with further processing and scaling using the CCP4 processing package (*23*). An initial solution to the Patterson function was produced in the tetragonal pseudo space group both automatically with the program SOLVE (*23*) and manually with the program RSPS (*23*), and initial phases were calculated with SHARP (*23*). Density modification including fourfold noncrystallographic (NCS) averaging was done with the program DM (*23*). A model corresponding to the two copies of the complex in the pseudo asymmetric unit was built with the program O (*23*). Bulk solvent correction, positional, simulated annealing, and isotropic temperature factor refinement was carried out with X-PLOR (*23*) and REFMAC (*23*) with tight NCS restraints against a 2.5 Å native data set collected at Max II/BL711 in Lund. The current R factor and R_{free} (on 5% of the data) are 24.0 and 26.8%, respectively. The root mean square deviations from ideal bond lengths and angles are 0.016 and 3.3 Å, respectively. No residues are in disallowed regions of the Ramachandran plot.

Data collection statistics

Crystal	d_{min} (Å)	N_{unique}	Cmplt* (%)	Mult†	$I/\sigma(I)$‡	R_{sym}§ (%)	R_{anom}‖ (%)
SeMet Crystal 1	2.8		82.8				
Remote		93,019		2.5	13.1 (3.7)	4.0 (17.3)	3.5 (16.8)
Point of inflection		75,467		2.1	11.6 (6.9)	3.5 (24.4)	4.3 (21.4)
Peak		82,754		2.7	11.3 (1.9)	4.1 (24.7)	4.2 (18.8)
SeMet Crystal 2	2.7		98.7				
Remote		110,928		3.8	8.9 (2.0)	5.1 (28.3)	4.2 (20.9)
Point of inflection		110,415		4.0	10.6 (2.7)	4.2 (21.8)	3.8 (17.4)
Peak		110,418		3.9	14.4 (2.8)	4.2 (20.8)	4.2 (17.5)
Native	2.5	139,645	98.0	4.1	5.3 (1.6)	7.6 (25.3)	NA

Phasing statistics from SHARP

	Point of inflexion λ = 0.9793 Å		Peak λ = 0.9792 Å		Remote λ = 0.885 Å	
	Centric	Acentric	Centric	Acentric	Centric	Acentric
Phasing power¶	2.0/--	2.1/1.2	2.0/--	2.0/1.6	--/--	--/0.81
R_{cullis}#	0.49/--	0.56/0.52	0.53/--	0.54/0.57	--/--	--/0.69
Resolution (Å)	7.59 5.50	4.52	3.93	3.53	3.23	2.99	2.80
FOM**	0.623 0.508	0.379	0.227	0.172	0.140	0.105	0.125

*Completeness †Multiplicity ‡Overall value and values in parentheses are for the highest resolution shell. §$R_{sym} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ and $\langle I(h)\rangle$ are the intensities of the individual and mean structure factors, respectively. The high-resolution shell is in parentheses. ‖$R_{anom} = \Sigma_h|\Sigma_i I_i(h)^+ - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, $I_i(h)^+$ and $\langle I(h)\rangle$ are as defined above, and the summation is over anomalous pairs. The high-resolution shell is in parentheses. ¶F_H(calc)/E, where E is the estimated lack-of-closure error (isomorphous/anomalous). #$R_{cullis} = \Sigma||F_{PH} \pm F_P| - F_H(calc)|/\Sigma|F_{PH} - F_P|$, where F_P and F_{PH} are protein and heavy-atom structure factors, respectively, and F_H(calc) is the calculated heavy-atom structure factor (isomorphous/anomalous). **Figure of merit for SHARP phases.

[1]Department of Molecular Biology, Uppsala Biomedical Center, Swedish University of Agricultural Sciences, Box 590, S-751 24 Uppsala, Sweden. [2]European Molecular Biology Laboratory Grenoble Outstation, c/o Avenue des Martyrs, BP 156X, 38042 Grenoble, France. [3]European Synchrotron Radiation Facility, Avenue des Martyrs, 38400 Grenoble, France. [4]Medimmune, Gaithersburg, MD 20878, USA. [5]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA.

*To whom correspondence should be addressed. E-mails: hultgren@borstel.wustl.edu (S.J.H.); stefan@xray.bmc.uu.se (S.D.K.).

periplasm by forming chaperone-subunit complexes (*5*). Here, the x-ray crystal structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *E. coli* is described. The structure reveals a donor strand complementation mechanism that explains the basis of both chaperone function and pilus biogenesis.

The structure of the FimC-FimH complex was solved by means of multiwavelength anomalous dispersion (MAD) data to 2.7 Å collected from selenomethionyl FimC-FimH crystals, and subsequently refined to 2.5 Å (Table 1). Eight copies of the FimC-FimH heterodimer in the C2 asymmetric unit were arranged as two sets of four molecules related by approximate 4_1 screw axes. Electron density was excellent for one set of molecules (Fig. 1A), allowing us to trace the entire complex. For the second set of molecules, electron density was poorer but allowed for unambiguous placement of a copy of the initially traced complex.

FimH is folded into two domains of the all-beta class connected by a short extended linker (Fig. 1B). The NH_2-terminal mannose-binding lectin domain comprises residues 1H to 156H, and the COOH-terminal pilin domain, which is used to anchor the adhesin to the pilus, comprises residues 160H to 279H (Fig. 2A). The overall structure of the FimC chaperone in the complex is essentially the same as that of the free chaperone (*8*, *9*). The pilin domain of FimH binds in the cleft of the chaperone (Fig. 1B), although there is only limited contact between FimH and the COOH-terminal domain of FimC.

The lectin domain of FimH is an 11-stranded elongated β barrel with a jelly roll–like topology (Fig. 2B). Searches of the structural database (*10*, *11*) did not reveal any significant structural homologs of this domain. The fold starts with a short β hairpin that is not part of the jelly roll. The final (11th) strand of the domain is inserted between the 3rd and 10th strands and thus breaks the jelly-roll topology. A pocket capable of accommodating a monomannose unit is located at the tip of the domain, distal from the connection to the pilin domain (Fig. 1B). A molecule of cyclohexylbutanoyl-*N*-hydroxyethyl-D-glucamide (C-HEGA) (*12*) is bound in this pocket (Fig. 3A). The glucamide moiety of C-HEGA is blocked at C1 and cannot form a pyranose, but is bent to approach the pyranose conformation. The C2, C3, C4, and C6 hydroxyl groups of C-HEGA are enclosed within the pocket, whereas the C5 hydroxyl and cyclohexylbutanoyl-*N*-hydroxyethyl groups point out from the pocket and are solvent exposed. Residues Asp^{54H}, Gln^{133H}, Asn^{135H}, Asp^{140H}, and the NH_2-terminal amino group of FimH (Fig. 3A) are hydrogen bonded to the glucamide moiety of C-HEGA. FimH from a urinary tract *E. coli* isolate that has a lysine instead of asparagine at position 135H produces type 1 pili but is unable to mediate mannose-sensitive hemagglutination of guinea pig erythrocytes (*13*). Also, a mutation at residue 136H has been reported to completely block mannose binding (*14*).

The pilin domain of FimH has the same immunoglobulin-like topology as the NH_2-terminal domain of periplasmic chaperones, ex-

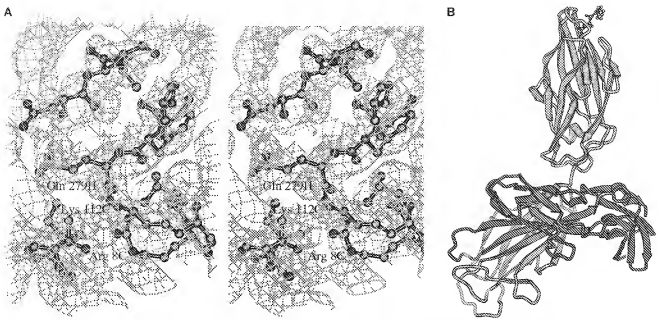

Fig. 1. (**A**) A typical sample of the solvent-flattened experimental electron density map (contoured at 1.0σ) with the refined model superimposed. Arg^{8C} and Lys^{112C} anchor the COOH-terminus of FimH ($Gln^{279†}$) in the subunit binding cleft of the chaperone through hydrogen bonds to the terminal carboxylate. (**B**) MOLSCRIPT (*24*) ribbon diagram of the FimC-FimH complex. FimH is colored yellow, except for the A″ (green) and F (orange) strands of the pilin domain. FimC is colored blue, except for the G1 strand, which is cyan. The FimH pilin domain and the NH_2-terminal domain of FimC form a closed superbarrel with a continuous core made from conserved residues in both proteins. A ball-and-stick representation of the C-HEGA molecule bound to the lectin domain of FimH indicates the position of the carbohydrate-binding site at the tip of the domain.

cept that the seventh strand of the fold is missing (Fig. 2B). Two antiparallel β sheets (strands A′BED′ and D″CF) pack against each other to form a β barrel that is similar to, but distinct from, immunoglobulin barrels. As in the chaperones, strand switching occurs at the edges of the sheets. In the chaperones, the A1 strand of the NH_2-terminal domain switches between the two sheets of the barrel (*15*). The first strand of the pilin domain exhibits a similar switch, but owing to the lack of a seventh strand, the second half of the A strand is not involved in main-chain hydrogen bonding within the domain. The D strand of the chaperones as well as of the FimH pilin domain also switches, but in the pilin domain the switch is an eight-residue loop instead of the cis-proline bulge found in the chaperones. The C-D loop and the D′-D″ connection pack against each other and close the top of the barrel. The other side of the barrel, defined by the A and F edge strands, is open. Owing to the absence of a seventh strand, a deep scar is created on the surface of the domain. Residues that would be part of the hydrophobic core of an intact, seven-stranded fold instead line a deep hydrophobic crevice on the surface of the pilin domain (Fig. 3B).

In the complex, the seventh (G1) strand from the NH_2-terminal domain of the chaperone is used to complement the pilin domain by being inserted between the second half of the A strand and the F strand of the domain (Fig. 3C). The final strand (F) of FimH forms a parallel β-strand interaction with the G1 strand of FimC and has its COOH-terminal carboxylate anchored at the bottom of the chaperone cleft

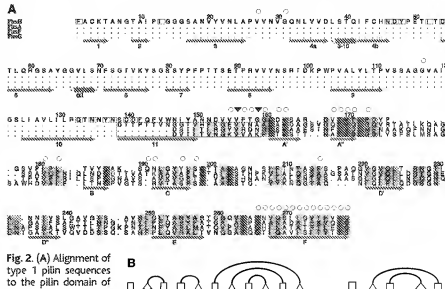

Fig. 2. (**A**) Alignment of type 1 pilin sequences to the pilin domain of FimH. The end of the lectin domain and the start of the pilin domain in FimH are indicated by black arrowheads above the sequences. Clustal W (*25*) was used to align the sequences, which were manually adjusted to minimize gaps (indicated by dots) in secondary structure elements. Residue 1 of FimH is residue 22 in the precursor protein (*26*). Residues are coded as follows: identical (red); conserved character (blue); pilin NH_2-terminal residues proposed to take part in donor strand complementation in the pilus (yellow); involved in chaperone binding (*27*) (open circle above the residue); carbohydrate binding pocket (boxed). The NH_2-terminal extensions of the pilin subunits are in one large box. Limits and nomenclature for secondary structure elements are shown below the sequence. (**B**) β-sheet topology diagrams of the lectin (left) and pilin (right) domains of FimH. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

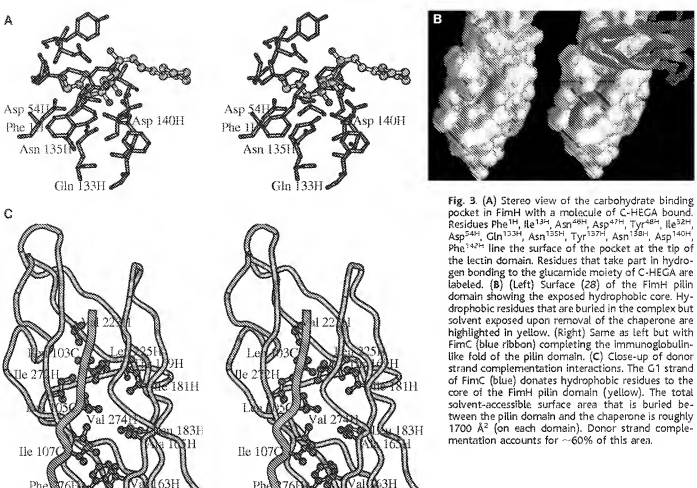

Fig. 3. (**A**) Stereo view of the carbohydrate binding pocket in FimH with a molecule of C-HEGA bound. Residues Phe^{1H}, Ile^{13H}, Asn^{46H}, Asp^{47H}, Tyr^{48H}, Ile^{52H}, Asp^{54H}, Gln^{133H}, Asn^{135H}, Tyr^{137H}, Asn^{138H}, Asp^{140H}, Phe^{142H} line the surface of the pocket at the tip of the lectin domain. Residues that take part in hydrogen bonding to the glucamide moiety of C-HEGA are labeled. (**B**) (Left) Surface (*28*) of the FimH pilin domain showing the exposed hydrophobic core. Hydrophobic residues that are buried in the complex but solvent exposed upon removal of the chaperone are highlighted in yellow. (Right) Same as left but with FimC (blue ribbon) completing the immunoglobulin-like fold of the pilin domain. (**C**) Close-up of donor strand complementation interactions. The G1 strand of FimC (blue) donates hydrophobic residues to the core of the FimH pilin domain (yellow). The total solvent-accessible surface area that is buried between the pilin domain and the chaperone is roughly 1700 Å^2 (on each domain). Donor strand complementation accounts for ~60% of this area.

through hydrogen bonding with the conserved residues Arg^{8C} and Lys^{112C} in FimC (Fig. 1A). This interaction is critical for chaperone function (*16, 17*).

The G1 strand of periplasmic chaperones contains a conserved motif of solvent-exposed hydrophobic residues at positions 103, 105, and 107 in FimC (*15*). In the complex, these residues are used to complete the unfinished hydrophobic core of FimH (Fig. 3C). The two residues Leu^{103C} and Leu^{105C} are deeply buried in the crevice created in the FimH pilin domain owing to the missing seventh strand, Ile^{107C} is somewhat closer to the domain surface but makes van der Waals contacts with residues Val^{163H} and Phe^{276H}. Leu^{103C} contacts residues Ile^{181H}, Val^{223H}, Leu^{325H}, and Ile^{272H}. Leu^{105C} is in contact with Ile^{181H}, Leu^{183H}, Leu^{252H}, Ile^{272H}, and Val^{274H}. We denote this mode of binding "donor strand complementation" to emphasize the fact that the pilin domain is incomplete and that the chaperone donates its G1 strand to complete the fold. Donor strand complementation has also been observed in the recent crystal structure of the PapD-PapK complex (*18*).

Genetic, biochemical, and electron microscopic studies have demonstrated that residues in two conserved motifs (the COOH-terminal F strand and an NH_2-terminal motif) participate in subunit-subunit interactions necessary for pilus assembly (*17*). An alignment of the pilin sequences, based on the FimC-FimH crystal structure, revealed that the NH_2-terminal motif was part of a 10- to 20-residue NH_2-terminal extension that was missing in the FimH pilin domain (Fig. 2A) and disordered in the PapD-PapK complex (*18*). This region contains a pattern of alternating hydrophobic residues similar to the G1 donor strand of the chaperone. On the basis of molecular modeling, the NH_2-terminal extension of a subunit is predicted to be able to take the place of the G1 strand of the chaperone, and fit into the pilin groove. Thus, during pilus assembly, alternating hydrophobic side chains in the NH_2-terminal extension could replace the hydrophobic side chains donated to the pilin core by the G1 strand of the chaperone, through a donor strand exchange mechanism. Thus, every subunit would complete the immunoglobulin-like fold of its neighboring subunit.

The type 1 pilus is a right-handed helix with about three subunits per turn, a diameter of ~70 Å, a central pore of about 20 to 25 Å, and a pitch of about 24 Å (*19*). To obtain this structure, insertion of the NH_2-terminal extension must be antiparallel to strand F, in contrast to the parallel insertion observed for the G1 strand of the chaperone. Insertion in a parallel orientation would lead to rosettelike structures. Using the FimH pilin domain as a model for FimA, we constructed a model for the type 1 pilus that fit these data (Fig. 4). Each subunit was aligned to have its cleft facing toward the center of the pilus so that the height from the top to the bottom of the domain along the helix axis was ~25 Å. By applying a rotation of 115° and a rise per subunit of 8 Å, a hollow helical cylinder is created. The outer diameter of this cylinder as measured across Cα atoms is 70 Å, and the inner diameter is 25 Å. FimA subunits from

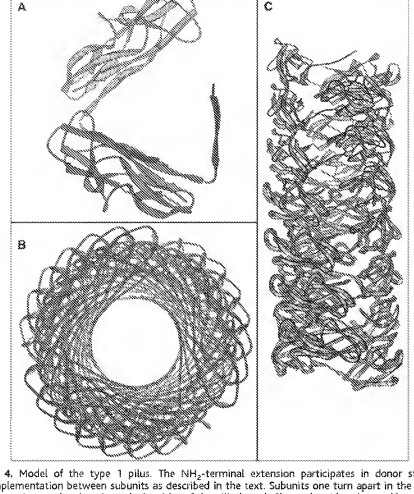

Fig. 4. Model of the type 1 pilus. The NH_2-terminal extension participates in donor strand complementation between subunits as described in the text. Subunits one turn apart in the helix pack against each other through the sides of the pilin barrel. Charged residues located between the hydrophobic side chains in the NH_2-terminal extension point into the solution on the inside of the hollow pilus rod. **(A)** The proposed interaction between two consecutive FimA molecules in the type 1 pilus rod. The modeled NH_2-terminal extension is colored red. **(B)** View of the pilus from the top. Residue positions that are subject to allelic variation (shown in blue) map to the outer surface of the pilus. **(C)** Side view of the pilus.

different strains of *E. coli* exhibit considerable allelic variation (*13*). The vast majority of the variable positions are on the outside surface of the pilus model proposed above (Fig. 4), which would account for the antigenic variability of type 1 pili.

The proposed head-to-tail interaction between subunits in a pilus is reminiscent of oligomerization through 3D domain swapping (*20*), in the sense that a part of one protein molecule is used to complement another. However, in this case, complementation occurs not only between identical protein chains (FimA in the pilus rod), but also between homologous but distinct chains (for example, FimG. FimF, and FimH in the pilus tip). Furthermore, because individual pilin protomers do not exist as stable monomers, there is no exchange of structural units between a monomeric and an oligomeric state. Instead, a different protein, the periplasmic chaperone, is needed to keep the monomeric subunits in solution by donating a unique part of its structure (the G1 strand) to the different subunit grooves.

On the basis of the structure of the FimC-FimH complex, we propose that the class of proteins known as pilins are missing necessary steric information needed to fold into a native three-dimensional structure. The information that is missing consists of the seventh-edge strand of an immunoglobulin fold. This strand, which is necessary for folding, is donated to the hydrophobic core of the pilin by the periplasmic chaperone in a donor strand complementation mechanism. A recent formulation of Anfinsen's classic postulate stated that "The steric information necessary for newly synthesized protein chains to fold correctly within cells resides solely in the primary structure of the initial translation product" (*21*). Here we provide an example of a case where some of that information is not inherent in the sequence of the protein to be folded but is instead transferred from another protein—the periplasmic chaperone.

References and Notes

1. C. H. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2081 (1995).
2. S. N. Abraham, D. Sun, J. B. Dale, E. H. Beachey, *Nature* **336**, 682 (1988).
3. K. A. Krogfelt, H. Bergmans, P. Klemm, *Infect. Immun.* **58**, 1995 (1990).
4. M. A. Mulvey *et al.*, *Science* **282**, 1494 (1998).
5. C. E. Soto and S. J. Hultgren, *J. Bacteriol.* **181**, 1059 (1999).
6. D. G. Thanassi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3146 (1998).
7. A. Holmgren and C.-I. Brändén, *Nature* **342**, 248 (1989).
8. M. Pellecchia, P. Guntert, R. Glockshuber, K. Wuthrich, *Nature Struct. Biol.* **5**, 885 (1998).
9. S. D. Knight, unpublished results.
10. G. J. Kleywegt and T. A. Jones, *Methods Enzymol.* **277**, 525 (1997).
11. L. Holm and C. Sander, *J. Mol. Biol.* **233**, 123 (1993).
12. C-HEGA is not a known inhibitor of FimH mannose binding but was needed in the crystallization [S. D. Knight, M. Mulvey, J. Pinkner, *Acta Crystallogr.* **D53**, 207 (1997)] to produce useful crystals. Briefly, FimC-FimH crystals were grown by hanging drop vapor diffusion by mixing 2 μl of a protein solution (4 mg of FimC-FimH per milliliter pre-equilibrated in 300 mM C-HEGA) with 2 μl of reservoir solution containing 1 M ammonium sulfate in 0.1 M tris-HCl buffer (pH 8.2).
13. S. Langermann, unpublished results.
14. M. A. Schembri, L. Pallesen, H. Connell, D. L. Hasty, P. Klemm, *FEMS Microbiol. Lett.* **137**, 257 (1996).
15. D. L. Hung, S. D. Knight, R. M. Woods, J. S. Pinkner, S. J. Hultgren, *EMBO J.* **15**, 3792 (1996).
16. M. J. Kuehn *et al.*, *Science* **262**, 1234 (1993).
17. G. F. Soto *et al.*, *EMBO J.* **17**, 6155 (1998).
18. F. G. Sauer, G. Waksman, J. Pinkner, K. Futterer, K. W. Dodson, *Science* **285**, 1058 (1999).
19. C. C. Brinton Jr., *Trans. N.Y. Acad. Sci.* **27**, 1003 (1965).
20. M. P. Schlunegger, M. J. Bennett, D. Eisenberg, *Adv. Protein Chem.* **50**, 61 (1997).
21. R. J. Ellis and F. U. Hartl, *Curr. Opin. Struct Biol.* **9**, 102 (1999).
22. W. A. Hendrickson, *Science* **254**, 51 (1991).
23. Crystallographic software: HKL2000 [Z. Otwinowski and W. Minor, *Methods Enzymol.* **276**, 307 (1997)]; CCP4 processing package [CCP4, *Acta Crystallogr.* **D50**, 760 (1994)]; SOLVE [T. C. Terwilliger and J. Berendzen, *ibid.* **D53**, 571 (1997)]; RSPS [S. D. Knight, L. Andersson, C.-I. Brändén, *J. Mol. Biol.* **215**, 113 (1990)]; SHARP [E. de la Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997)]; DM [K. D. Cowtan, *Joint CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34 (1994)]; O [T. A. Jones, J.-Y. Zou, S. W. Cowtan, M. Kjeldgard, *Acta Crystallogr.* **A47**, 110 (1991)]; X-PLOR [A. T. Brünger, *X-PLOR Manual (Version 3.1): A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992)]; REFMAC [G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr.* **D53**, 240 (1997)].
24. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
25. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
26. Pilus subunits are expressed in the cytoplasm as precursor proteins with an NH_2-terminal signal sequence that is cleaved during transport across the inner membrane by the Sec machinery [S. J. Hultgren, S. Normark, S. N. Abraham, *Annu. Rev. Microbiol.* **45**, 383 (1991)]. The first visible FimH residue in our maps corresponds to Phe[1] in the gene-derived sequence, which is the expected start of the mature FimH chain [M. S. Hanson, J. Hemphel, C. B. Brinton Jr.,

J. Bacteriol. **170**, 3350 (1988)]. To distinguish residues in the adhesin from residues in the chaperone, FimH residues will be denoted by an H, and FimC residues by a C, after the residue number.
27. Interface residues were defined as having a difference in solvent accessibility [S. Miller, J. Janin, A. M. Lesk, C. Chothia, *J. Mol. Biol.* **196**, 641 (1987)] between the subunit in the complex and removed from the complex exceeding 10 percentage points
28. A. Nicholls, *GRASP, Graphical Representation and Analysis of Surface Properties* (Columbia Univ. Press, New York, 1993).
29. We thank A. Revel and J. Burlein for technical assistance and advice; the staff at the Max II synchrotron in Lund: H. Eklund, J. Hajdu, A. Jones, and S. Ramaswamy for discussions and reading of the manuscript; and J. Berglund for help with the figures. Supported by grants from the Swedish Research Council NFR and the Swedish Foundation for Strategic Research (Structural Biology Network) (S.D.A.), and by National Institutes of Health grants RO1DK51406 and RO1AI29549 (S.J.H.). The coordinates have been deposited at the Research Collaboratory for Structural Bioinformatics Protein Data Bank (code 1QUN).

25 March 1999; accepted 14 July 1999

Requirement of Circadian Genes for Cocaine Sensitization in *Drosophila*

Rozi Andretic, Sarah Chaney, Jay Hirsh*

The circadian clock consists of a feedback loop in which clock genes are rhythmically expressed, giving rise to cycling levels of RNA and proteins. Four of the five circadian genes identified to date influence responsiveness to free-base cocaine in the fruit fly, *Drosophila melanogaster*. Sensitization to repeated cocaine exposures, a phenomenon also seen in humans and animal models and associated with enhanced drug craving, is eliminated in flies mutant for *period*, *clock*, *cycle*, and *doubletime*, but not in flies lacking the gene *timeless*. Flies that do not sensitize owing to lack of these genes do not show the induction of tyrosine decarboxylase normally seen after cocaine exposure. These findings indicate unexpected roles for these genes in regulating cocaine sensitization and indicate that they function as regulators of tyrosine decarboxylase.

In response to exposure to volatilized free-base cocaine, *Drosophila* perform a set of reflexive behaviors similar to those observed in vertebrate animals, including grooming, proboscis extension, and unusual circling locomotor behaviors (*1*–*3*). Additionally, flies can show sensitization after even a single exposure to cocaine provided that the doses are separated by an interval of 6 to 24 hours (*1*). Sensitization, a process in which repeated exposure to low doses of a drug leads to increased severity of responses, has been linked to the addictive process in humans (*4*–*6*) and is potentially involved in the enhanced craving and psychoses that occur after repeated psychostimulant administration.

We have shown circadian variation in the agonist responsiveness of *Drosophila* nerve cord dopamine receptors functionally coupled to locomotor output (*7*). This variation is dependent on the normal functioning of the *Drosophila period* (*per*) gene, the founding member of the circadian gene family (*8*, *9*). Because changes in postsynaptic dopamine receptor responsiveness are also seen during cocaine sensitization in vertebrates (*10*–*12*), we examined flies mutant in circadian functions for alterations in responsiveness to cocaine.

Wild-type (WT) flies or flies containing a *per* null mutation, *per*^{o}, were exposed to 75 μg of cocaine four times over 2 days, and the fraction of flies showing severe responses was quantified after each exposure (Fig. 1A). Whereas WT flies showed sensitization after the initial cocaine exposure, *per*^{o} flies showed no sensitization either to a normal or increased dose even after repeated exposures. As with WT flies, *per*^{o} flies showed a dose-dependent increase in the severity of responses, and the normal cocaine-induced types of behaviors were observed (*13*).

per alleles that either shorten or lengthen the circadian periods show distinct patterns of cocaine responsiveness. The short-period mutants *per*^{s} and *per*^{T} (*14*, *15*) both showed increased responsiveness to the initial cocaine exposure and weak sensitization to a second 75-μg exposure (Fig. 2A), with only the sensitization of *per*^{s} showing statistical significance. Sensitization is not observed in these lines when tested with other cocaine doses (*16*). The long-period mutant *per*^{L1} (*17*) showed a normal initial cocaine response but no sensitization to a subsequent exposure.

Similarly, other circadian genes showed effects on cocaine sensitization: Both *clock* and *cycle* mutants failed to sensitize when given two doses of cocaine (Fig. 2B). Because these mutants showed an increased sensitivity to the

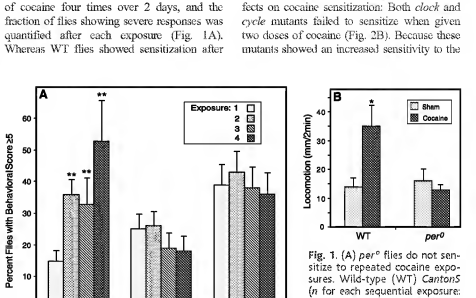

Fig. 1. (A) *per*^{o} flies do not sensitize to repeated cocaine exposures. Wild-type (WT) *Canton*S (*n* for each sequential exposure: 105, 95, 30, 17); and *per*^{o} (*n*: 81, 60, 61, 57) males were exposed to either 75 or 100 μg of volatilized free-base cocaine twice per day at 6-hour intervals for 2 days, and the behavioral responses were scored during the 5 min after exposure with a behavioral scale (*1*). Behavioral scores range from 0 (normal behavior) to 7 (death). Behavioral scores of ≥5 indicate rapid twirling, erratic jumping, or paralysis. Significant differences in responses to the first versus subsequent exposures (χ^2 test): $^{**}P \leq 0.01$. Error bars are standard deviations calculated for binomial distributions. All behavioral analyses were performed blindly; strains were given drugs in random order by placing a numbered tag in the video field during videotaping. The evaluator assayed behaviors blindly, and was unblinded only after all flies had been scored. (B) *per*^{o} flies do not modulate quinpirole responsiveness after cocaine exposures. WT *Oregon*^{R} and *per*^{o} were exposed to 75 μg of volatilized cocaine three times over 2 days and decapitated 4 hours after the last exposure. Flies were decapitated and assayed for locomotion with 2 mM quinpirole as described (*24*), with modifications (*7*, *31*). Average locomotion ± SEM is shown (n = 30 to 50 flies). Significant differences between sham- and cocaine-treated flies are indicated ($^{*}P \leq 0.05$, Student's t test).

Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903, USA.

*To whom correspondence should be addressed. E-mail: jh6x@virginia.edu